

LA-UR-18-26973

Approved for public release; distribution is unlimited.

Title: Quantifying Uncertainty in the Space Weather Forecasting

Author(s): Chakraborty, Shibaji
Morley, Steven Karl

Intended for: Internal Presentation to ISR-1 group

Issued: 2018-07-26

Disclaimer:

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.



Quantifying Uncertainty in Space Weather Forecasting

Probabilistic K_p Forecasting Model

Shibaji Chakraborty, Steven Karl Morley

ISR-1 Space Weather Summer School, 2018

UNCLASSIFIED

About Me

- *Shibaji Chakraborty*
- 3rd Year Ph.D. student at *SuperDARN group, Virginia Tech.*
- Working with *Dr. J. Michael Ruohoniemi & Dr. Joseph Baker.*
- Research interest is effects & impacts of space weather on trans-ionospheric HF communication.

UNCLASSIFIED

Outline

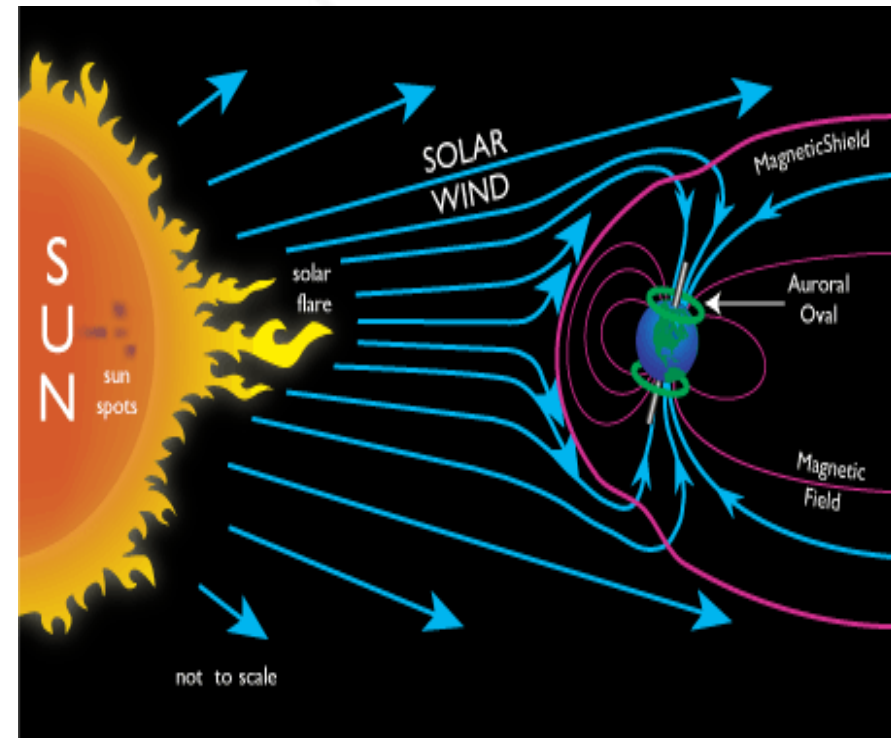
- Introduction & Motivation
- Objectives
- K_p : 3h Range Index & Current State of the Art for K_p Forecasting
- Forecasting K_p using machine learning
 - Nonparametric-Bayesian approach for Uncertainty Quantification
 - Comparing Different Dynamic Linear Models:
 - GLM & Ensemble Models
 - Gaussian Process Regression.
 - Deep Gaussian Process.
- Probabilistic Storm Forecast
- Conclusion & Future Work

UNCLASSIFIED

Slide 3

Introduction and Motivation

- Sun is the source of geomagnetic storms.
- Consequences of these solar wind driven storms can cost billions of dollars
- Hence forecasting space weather is a major challenge addressing the security of modern technology.
- In this study aims to provide a new method to forecast K_p with uncertainty bound associated with each prediction.



UNCLASSIFIED

Slide 4

Objectives

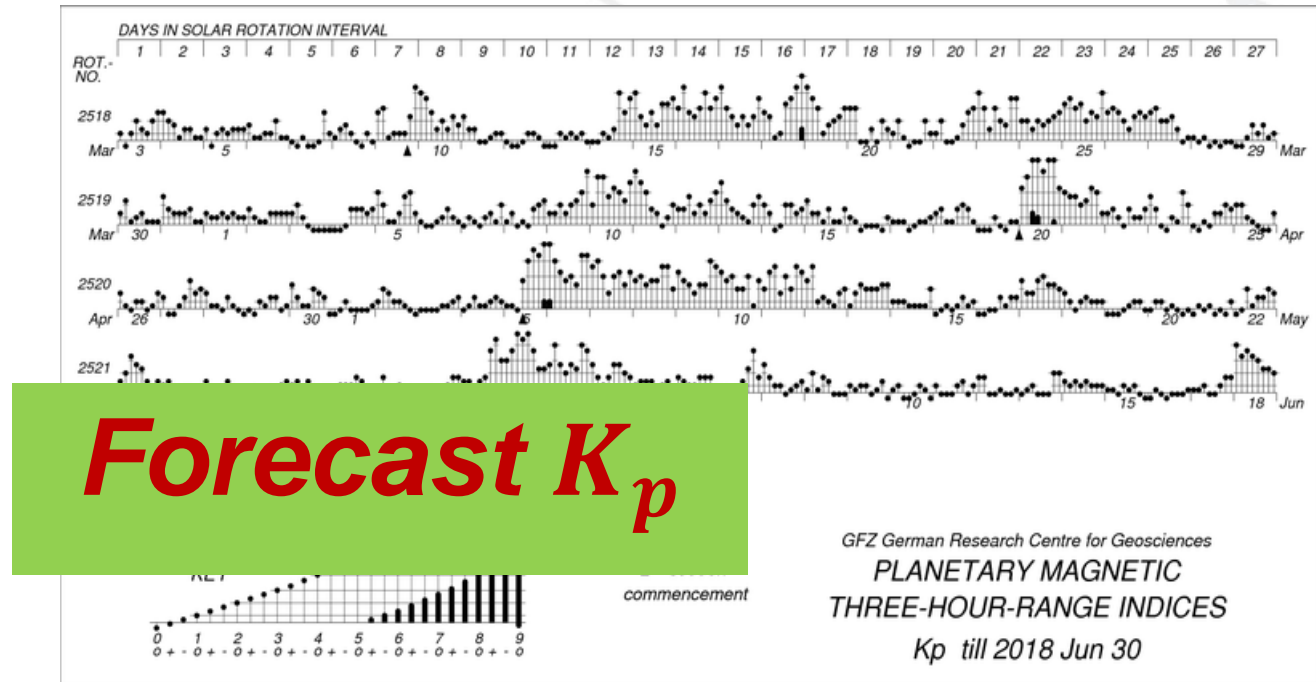
- **Provide a probabilistic geomagnetic storm forecast model.**
- **Provide an uncertainty quantification associated with prediction.**
- Provide insights about the solar wind parameters and solar cycle and how they affect the coupling to the geospace environment.

UNCLASSIFIED

K_p : 3h Range Index

What is K_p ?

- a. K_p is a 3 hourly planetary *range* index.
- b. Presents disturbances in the of the Earth's magnetic field.
- c. Integer in the range 0–9.



Importance of K_p in forecasting storms:

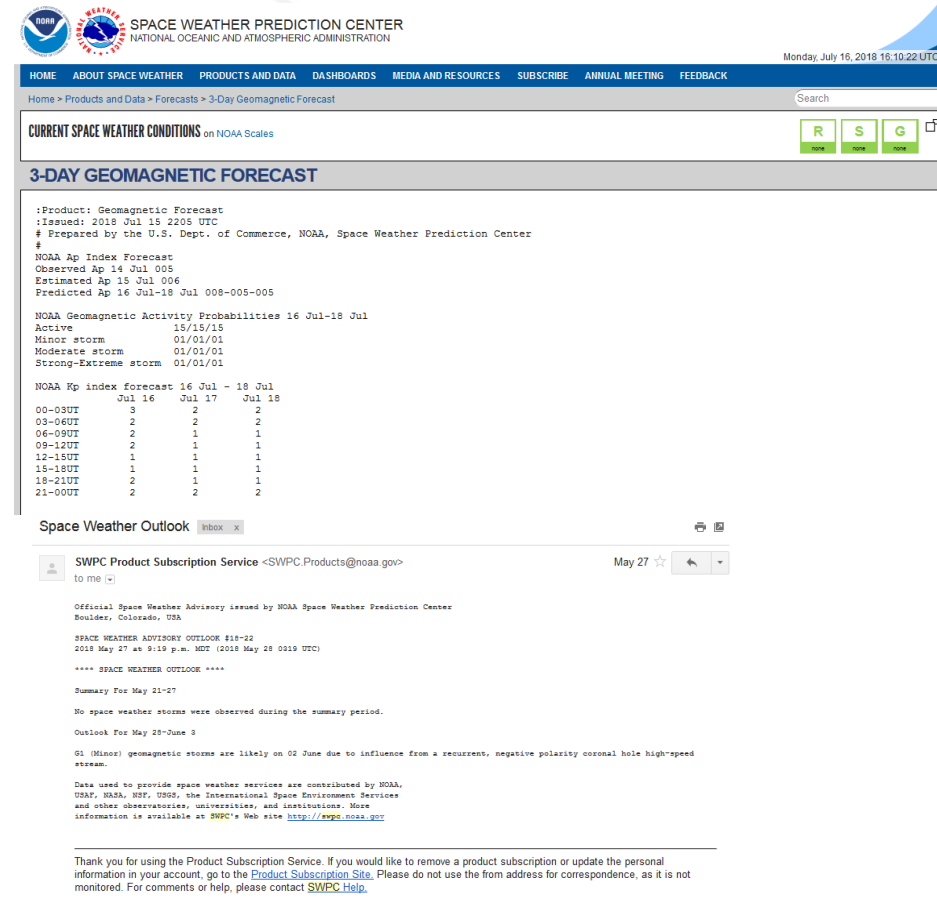
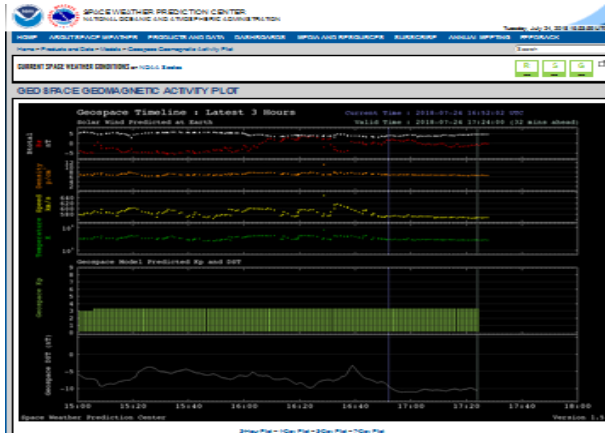
Used by **SWPC** to decide geomagnetic alerts and warnings need to be issued for users who are affected by these disturbances.

UNCLASSIFIED

Slide 6

Current State of the Art for K_p Forecasting

1. K_p forecast models, S. Wing et. al., J. Geophys. Res. [Currently operational at NOAA SWPC]



Crux: Existing models forecasts K_p either from empirically-derived coupling functions to forecast short-term K_p , or neural networks for 3-to-12 hour ahead prediction. But none of them provides an **Uncertainty** associated with predicted K_p .

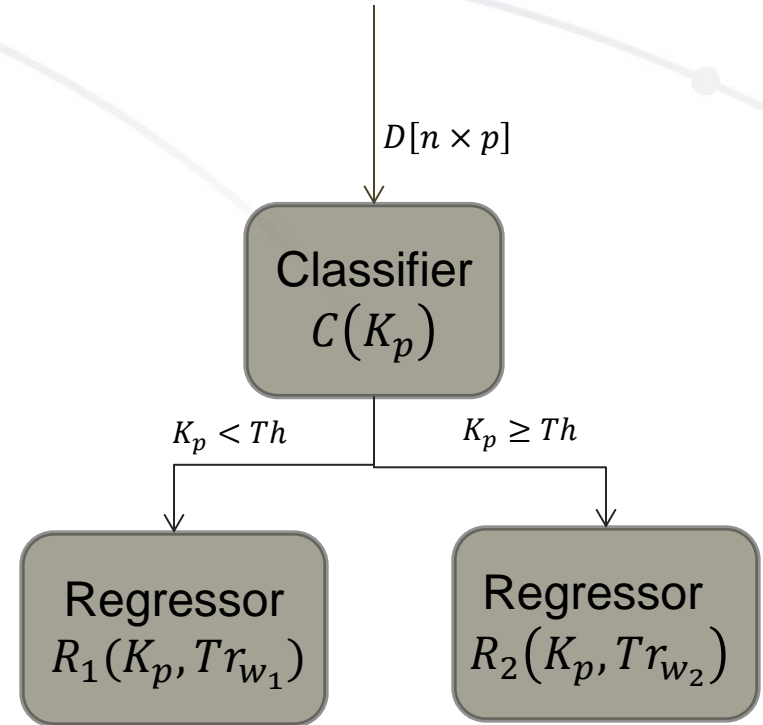
NOAA SWPC K_p and storm Forecast system

UNCLASSIFIED

Proposed Model

Characteristic:

- Subdivide the problem into two categories ($K_p < 5^-$ and $K_p \geq 5^-$).
- **Dynamic linear model instead of train model once with fixed dataset.**
- Classifier: Ensemble [Deterministic]
- Regressors: Deterministic (GLM, LSTM), Probabilistic (Gaussian Process Regression, deep GPR)



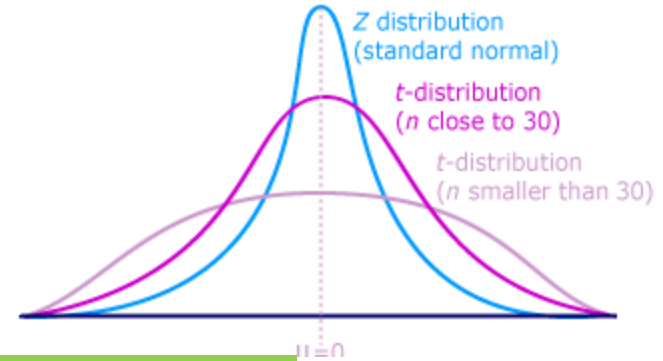
To get a probabilistic forecast with uncertainty bound we are going to use *Nonparametric-Bayesian Methods*

UNCLASSIFIED

Nonparametric Bayesian Methods

1. Parametric:

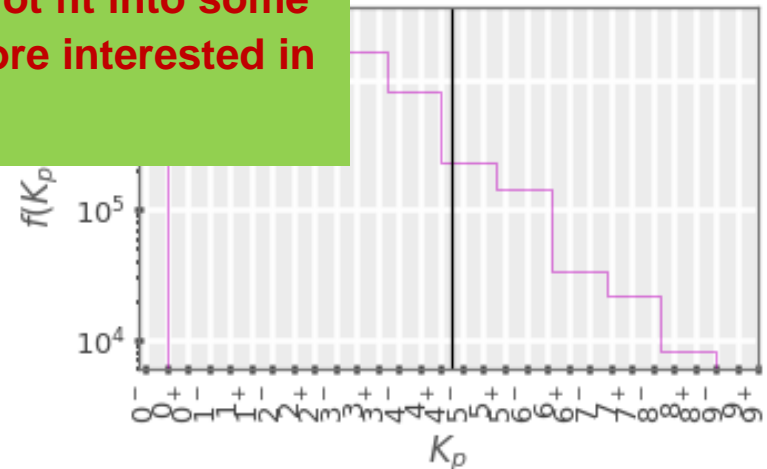
- Sample data comes from a population that follows a pdf based on a **fixed set of parameters** – *Normal, Beta, Uniform* ...
- Assumptions may lead to fitting errors.



2. Nonparametric:

- Nonparametric methods do not assume a specific distribution but with the distribution's parameters unspecified.
- Better **tail behavior**, outlier detection.

Clearly K_p has a heavy tail, does not fit into some known distribution, and we are more interested in forecasting the high K_p (outliers).



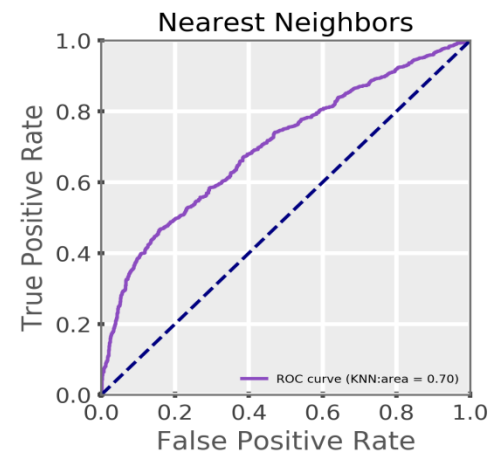
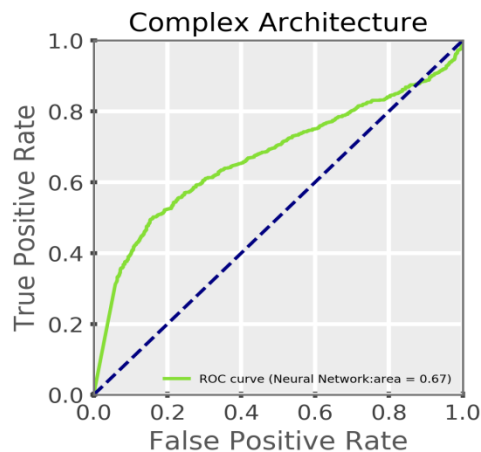
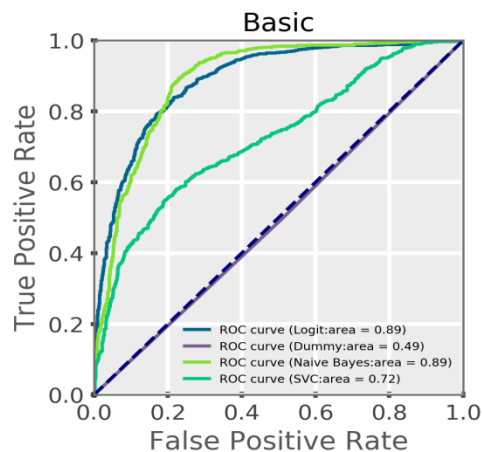
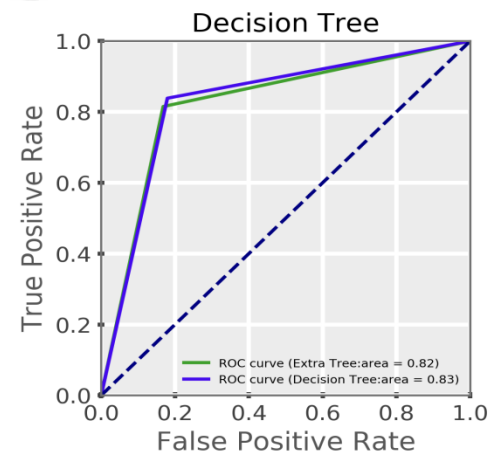
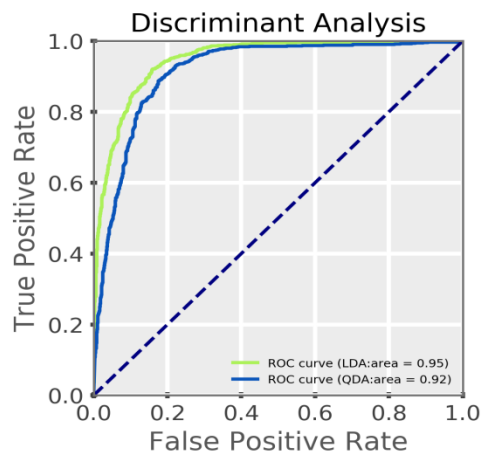
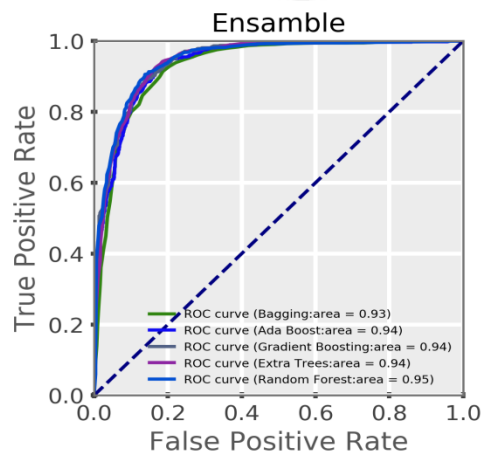
3. Nonparametric-Bayesian:

- $$\Pr(\theta|X) = \frac{\Pr(X|\theta) \Pr(\theta)}{\sum \Pr(X|\theta)}$$

UNCLASSIFIED

Slide 9

Classifier: Deterministic



UNCLASSIFIED

Slide 10

Regression: Gaussian Process (NB)

Gaussian Process Regression (Kriging):

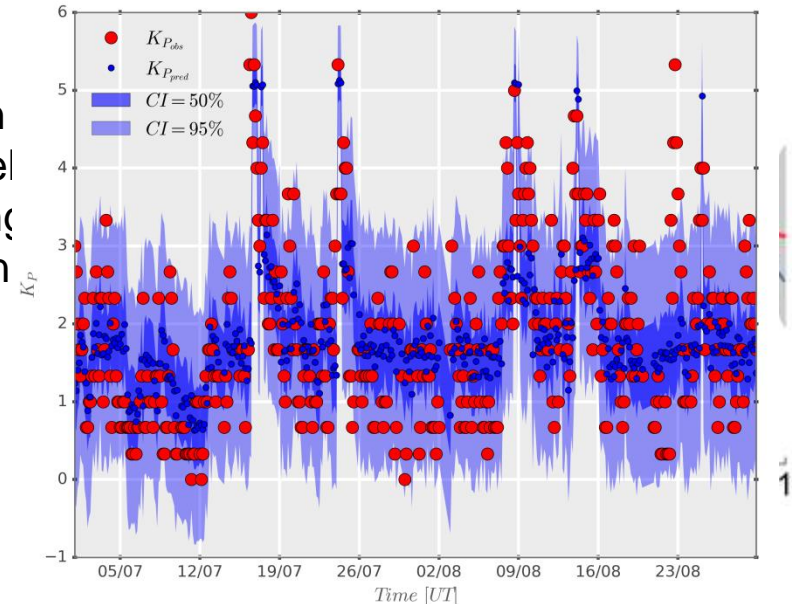
- GPR is another method (stochastic in nature) of interpolation where the model assumes the interpolated data is coming from the multivariate normal distribution

$$\begin{pmatrix} f \\ f_* \end{pmatrix} \sim N \left(\begin{pmatrix} m_X \\ m_{X_*} \end{pmatrix}, \begin{pmatrix} K_{XX} & K_{X_*X} \\ K_{X_*X} & K_{X_*X_*} \end{pmatrix} \right)$$

- Here K is the kernel function. It also assumes some prior about the kernel.

- $K_{SQ} = \sigma^2 e^{-\frac{(x-x')^2}{2l^2}}$

- Model takes multiple of 27 days training data to minimize RMSE during solar minima and takes 7-14 days of training data during solar maxima.

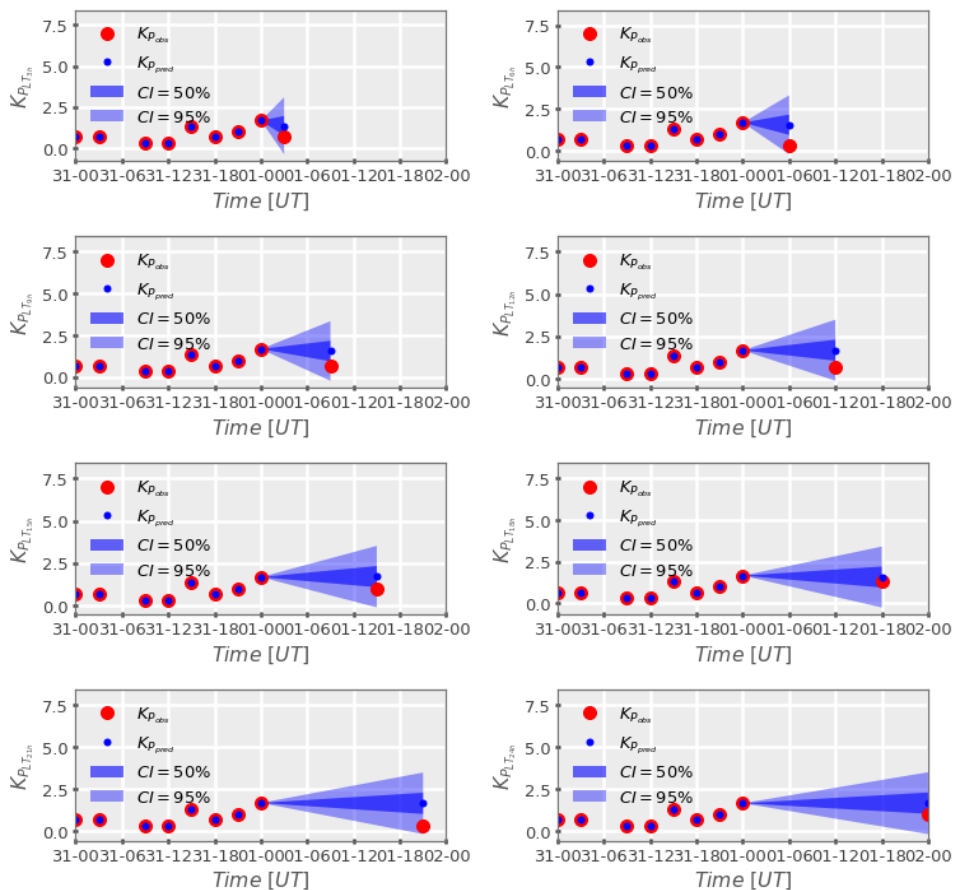


3h ahead of K_p prediction from July 1st 1995 – August 31st 1995

UNCLASSIFIED

Regression: Gaussian Process (Cont.)

How it can be used to get 24h forecast: [Still Working]



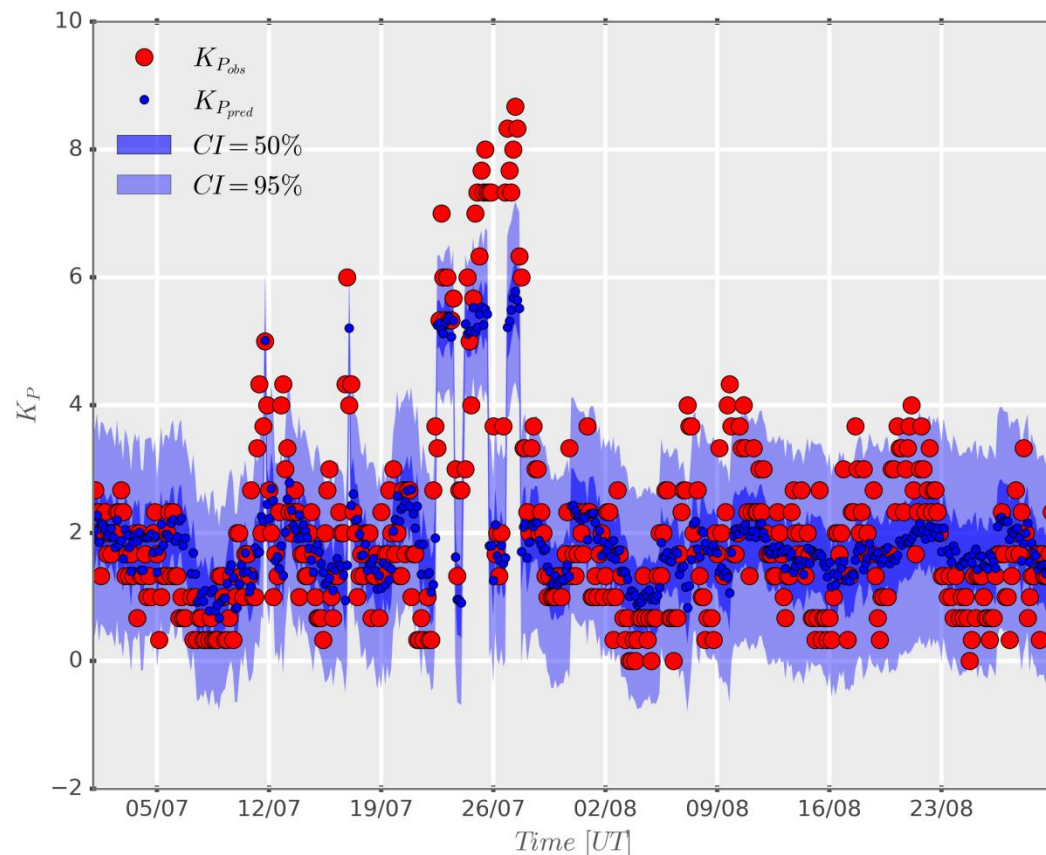
UNCLASSIFIED

Slide 12

Regression: Gaussian Process (Cont.)

Issues with GPR:

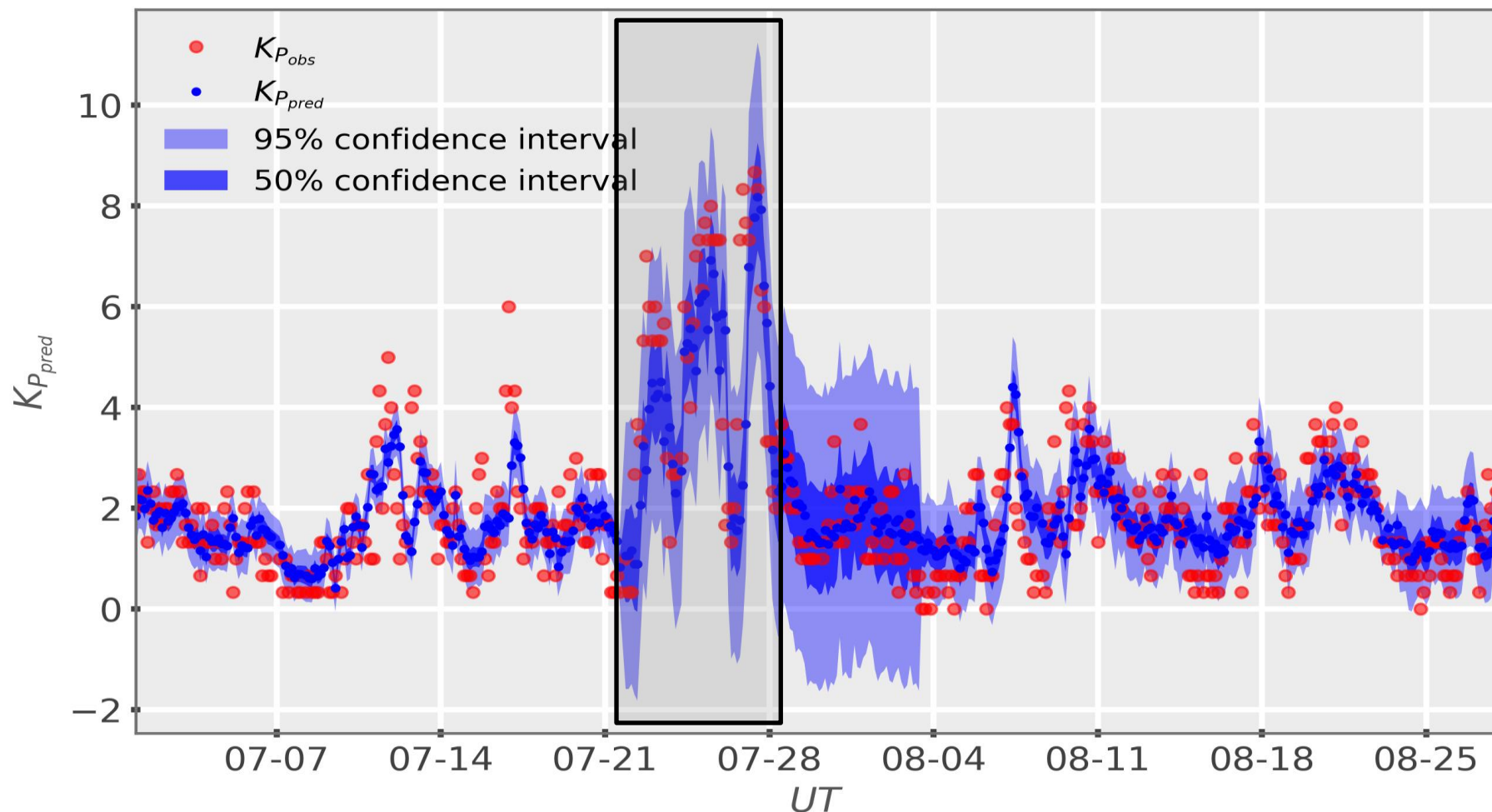
- Kernel selection
- Low accuracy
- **Not able to detect the transients in the system.**
[Not enough Physics]



UNCLASSIFIED

Slide 13

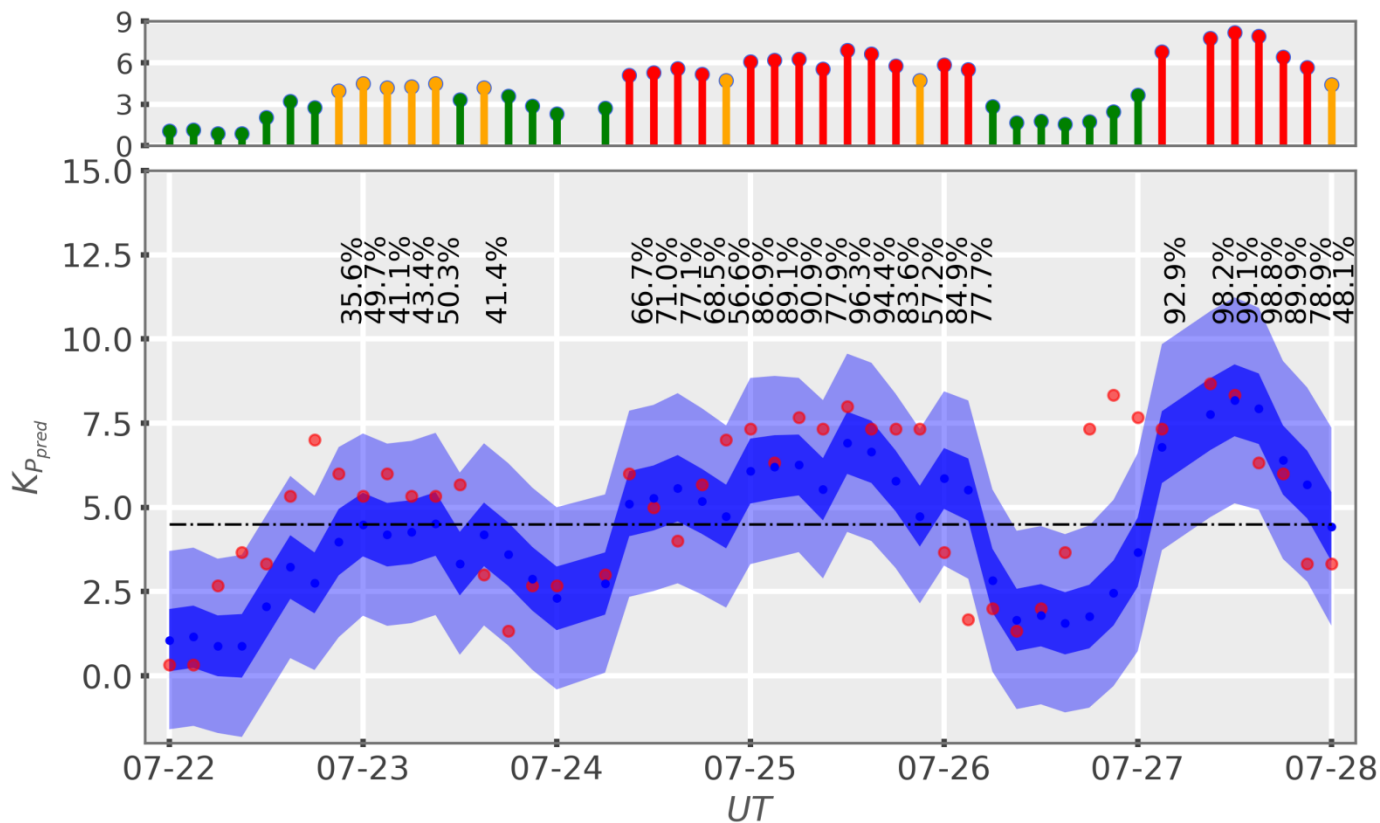
Deep GPR: RNN with Memory & GPR



UNCLASSIFIED

Slide 14

Probabilistic Storm Forecasting



Red: $\Pr(e \geq G_1) > 60\%$
 Orange: $30\% < \Pr(e \geq G_1) < 60\%$
 Green: $\Pr(e \geq G_1) < 30\%$

UNCLASSIFIED

Conclusions & Future Work

Conclusions:

1. Gaussian Processes with LSTM is a cutting edge tool for space weather forecast.
2. Dynamic linear model shows it takes 27 days to train a model during solar minima and 7-14 days for solar maxima.

Future Work:

1. *Run models for different conditions and different parameters.*
2. *Use different kernels for GP.*
3. *Introduce GOES X-ray & other dataset to capture more solar transients.*

UNCLASSIFIED

Slide 16



Questions
are
guaranteed in
life;
Answers
aren't.

Thank You!!

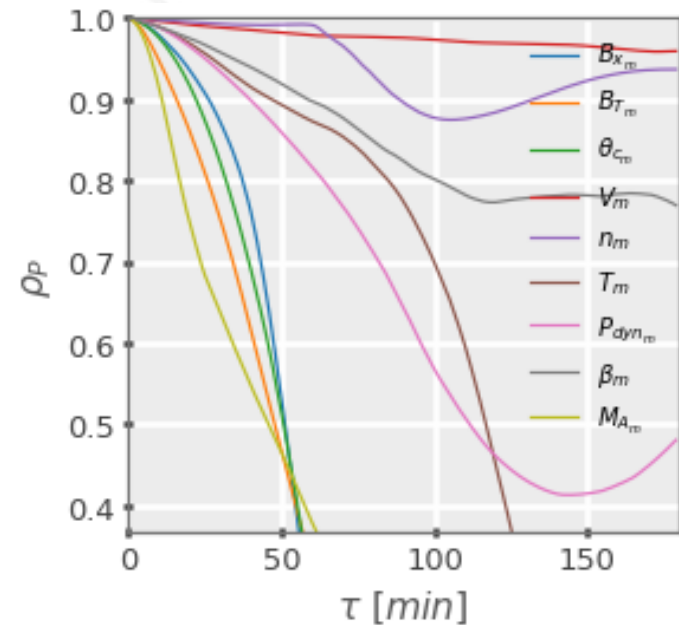
UNCLASSIFIED

Dataset: Preprocessing (Backup)

1. 1st Stage Data:

- Omni 1m resolution data and 3h K_p values.
- Used parameters: $B_x, B_t, \theta_c, v, n, T, P_{dyn}, \beta, M$
- We also used historical $K_{p_{hist}}$
- 3h average data.

2. Use GOES X-ray data:



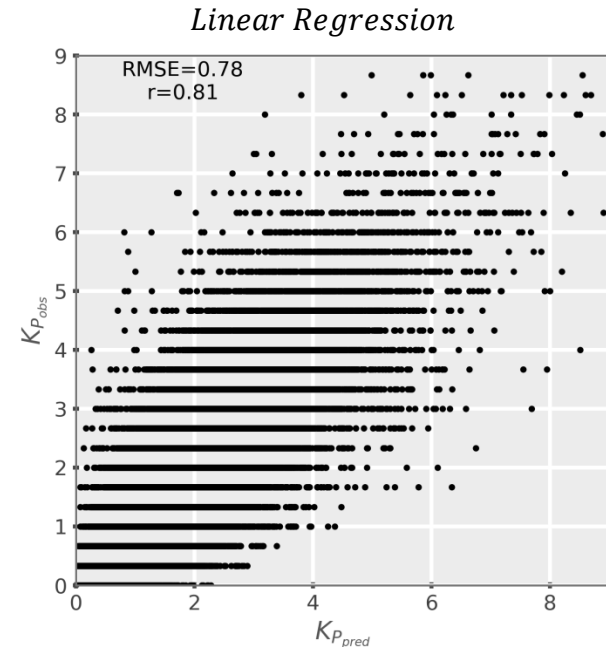
UNCLASSIFIED

Slide 18

Regression:

1. Algorithms:

- Basic: Linear Regression, Elastic Net, Bayesian Ridge
- Tree: Decision tree, Extra tree
- Ensemble: Bagging, Ada Boost, Random Forest
- Nearest Neighbors: KNN



Algorithms (RMSE, ρ)	$Tr_w = 27d$	$Tr_w = 54d$	$Tr_w = 81d$
Linear Regression	0.78,0.81	0.78,0.814	0.78,0.818
Elastic Net	0.85,0.775	0.84,0.80	0.85,0.774
Bayesian Ridge	0.78,0.81	0.78,0.814	0.78,0.818
D-tree	0.91,0.744	0.88,0.761	0.86,0.77
E-tree	0.91,0.744	0.88,0.761	0.86,0.77
Bagging	0.87,0.77	0.87,0.77	0.87,0.77
Ada Boosts	0.90,0.77	0.87,0.77	0.87,0.77
Random Forest	0.81,0.77	0.79,0.80	0.79,0.80
KNN	1.18,0.5	1.18,0.5	1.18,0.5

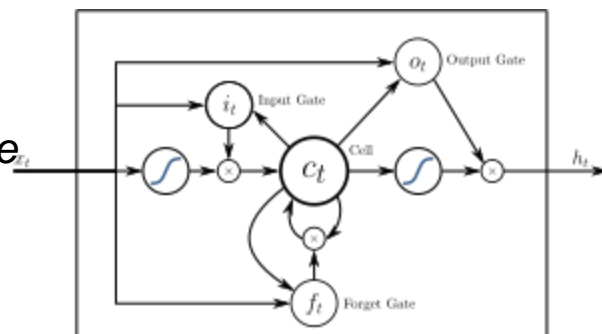
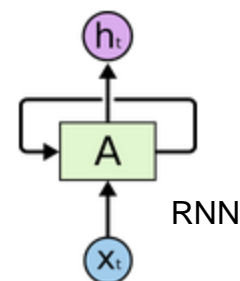
UNCLASSIFIED

Slide 19

Regression: LSTM (Backup)

1. LSTM: Long-short term memory (Recurrent Neural Network)

- LSTM units are a building unit for layers of a recurrent neural network.
- A common LSTM unit is composed of a **cell**, an **input gate**, an **output gate** and a **forget gate**.
- Cell is the memory unit.
- Input gate decides which values will be updated.
- Output gate decides which values will be updated at the output side.
- Forget gate discards a part or fully discard previous information.

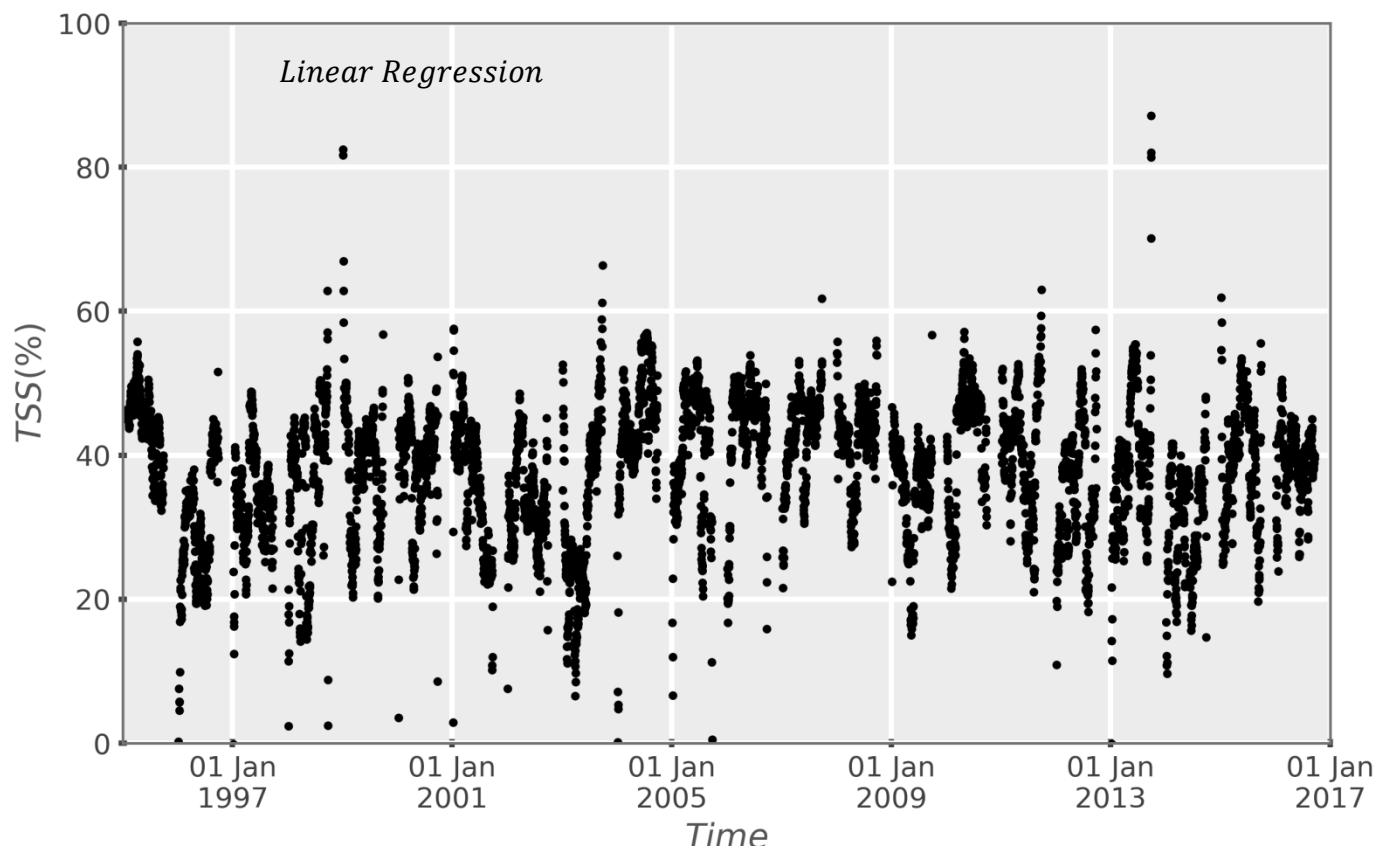


Algorithms (RMSE, ρ)	$Tr_w = 27d$	$Tr_w = 54d$	$Tr_w = 81d$
LSTM	0.94,0.73	—	—

UNCLASSIFIED

Slide 20

Model Evaluation: (Backup)



UNCLASSIFIED

Slide 21

Regression: Deep GPR (With GOES X-ray data)

1. We introduce GOES X-ray datasets:

- Try to capture the transients in the system.
- *Captures the solar magnetic activity as a proxy.*

Algorithms (RMSE, ρ)	$Tr_w = 27d$	$Tr_w = 54d$	$Tr_w = 81d$
Deep GPR (X-ray)	0.79,0.80	—	—

UNCLASSIFIED

Slide 22